COS 738: Assignment 1

Ryan Hansen

4025479

Task2

Cracking the substitution Cipher:

GFS WMY OG LGDVS MF SFNKYHOSU ESLLMRS, PC WS BFGW POL DMFRQMRS, PL OG CPFU M UPCCSKSFO HDMPFOSXO GC OIS LMES DMFRQMRS DGFR SFGQRI OG CPDD GFS LISSO GK LG, MFU OISF WS NGQFO OIS GNNQKKSFNSL GC SMNI DSOOSK. WS NMDD OIS EGLO CKSJQSFODY GNNQKKPFR DSOOSK OIS 'CPKLO', OIS FSXO EGLO GNNQKKPFR DSOOSK OIS 'LSNGFU' OIS CGDDGWPFR EGLO GNNQKKPFR DSOOSK OIS 'OIPKU', MFU LG GF, QFOPD WS MNNGQFO CGK MDD OIS UPCCSKSFO DSOOSKL PF OIS HDMPFOSXO LMEHDS. OISF WS DGGB MO OIS NPHISK OSXO WS WMFO OG LGDVS MFU WS MDLG NDMLLPCY POL LYEAGDL. WS CPFU OIS EGLO GNNQKKPFR LYEAGD MFU NIMFRS PO OG OIS CGKE GC OIS 'CPKLO' DSOOSK GC OIS HDMPFOSXO LMEHDS, OIS FSXO EGLO NGEEGF LYEAGD PL NIMFRSU OG OIS CGKE GC OIS 'LSNGFU' DSOOSK, MFU OIS CGDDGWPFR EGLO NGEEGF LYEAGD PL NIMFRSU OG OIS CGKE GC OIS 'OIPKU' DSOOSK, MFU LG GF, QFOPD WS MNNGQFO CGK MDD LYEAGDL GC OIS NKYHOGRKME WS WMFO OG LGDVS..

We will be using frequency analysis and our knowledge of the English Language to solve the cipher key, as well as tips from Cryptanalysis Hints.

First, we used an online tool to do a frequency analysis, to see oh often individual letters appear, the two most frequent letters in the cipher text are “S”(88) and “O”(85), and the most frequent appearing letters in the English language is “E” and “T”. We can make an educated guess and map “S”-> ”E”, and “O”->” T”.

We also see that “M” is a single letter word, and the only single letter words in the area “A” and “I”, so “M” is “A” or “I”.

Next, we will be looking at several Bigrams and Trigrams, to find common 2 or three-letter words, using the keys we already have and the patterns in their appearance. Note that when we particularly decipher the text, we’ll use “\_” for blank letters.

* “OIS”, appears 24 times in the cipher text. We know that “S”-> ”E”, and “O”->” T”, then “OIS -.> “T\_E”, the most common three letter word that follows that pattern of the letter is ‘THE’, SO WE MAP “I”->” H”.
* “OG”, appears 9 times in the cipher text. We know that “O”->” T”, then the most common matching two-letter word is “TO”, we can map “G” -> “O”.
* “GF” AND “GC”, appear 7 times in the cipher text. We know that “G”->” O”, then the most common matching two-letter word is “OF” and “ON”, we can map either “F” or “C” to “F” or “N”.
* “GFS”, appears 24 times in the cipher text. We know that “S”-> ”E”, and “G”->” O”, then “GFS” -.> “O\_E”, the most common three letter word that follows that pattern of the letter is ‘ONE’, so we can map “F”-> ”N”,
* and map “C” -> “F”
* “PF”, appears 12 times in the cipher text. We know that “F”-> “N”, then the most common matching two-letter word is “IN” OR “ON”, but we know “G” -> “O”, so we map “P”->” I” and can conclude that “M”-> “A”.
* “MFU”, appears 6 times in the cipher text. We know that “M”-> ”A”, and “F”->” N”, then “MFU” -.> “AN\_”, the most common three letter word that follows that pattern of the letter is ‘AND’, so we can map “U”-> ”D”.

AT THIS POINT WE HAVE

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CIPHER | A | B | C | D | E | F | G | H | I | J | K | L | M |
| PLAIN |  |  | F |  |  | N | O |  | H |  |  |  | A |
| CIPHER | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
| PLAIN |  | T | I |  |  | E |  | D |  |  |  |  |  |

At this point, we will be finding partial ciphers using the letter we already got and using our understanding of English and reading the words or sentence in context, to guess what the other keys map to.

* We examine CPDD, we know “C”-> “F” and “P” -> “I”. Gives us the partial text “FI\_ \_”. The list of most common doubles in English are “ss”, “ee”, “tt”, “mm”, “ll”, and “oo”. “D”-> “L” makes the most sense to give the word “CPDD”-> “FILL”.
* “PGFR” partial deciphers to “LONG\_”, then the most likely English word is “LONG” meaning “R”->” G”
* “DSOOSK” partially deciphers to “LETTE\_”, then the most likely English word is “LETTER”, then for “K”->” R”
* “DMFRQMRS” partially ciphers to “LANG\_AGE”, then the most likely English word is “LANGUAGE”, then for “Q”->” U”
* “ESLLMRS” partially deciphers to “ME\_ \_ AGE”, then using the list of common English doubles the most likely English word is “MESSAGE”, then for “L”->” S”
* Reading the first 4 words we have “GFS WMY OG LGDVS” partially ciphers to “ONE WAY TO SOLVE ”, WHERE “W”, “Y”, and “V” are still in cipher text but make sense in the context of the sentence. We, therefore map “W”->”W”, ”Y”-> “Y” and “V”-> “V”.
* “GNNUQKKPFR” partially deciphers to “O\_ \_URRN\_E”, and “MNNGQFO” deciphers to “A\_ \_ OUNT”, with the common letter “N”, then the most likely English words are “OCCURRENCE ” and “ACCOUNT”, then for “N”->” C”
* “ESLLMRS” partially deciphers to “ME\_ \_ AGE”, then using the list of common English doubles the most likely English word is “MESSAGE”, then for “L”->” S”
* Reading the line “OIS FSXO EGLO” partially ciphers to “THE NEXT MOST ”, WHERE “X” is still in cipher text but makes sense in the context of the sentence. We, therefore map “X”-> “X”.
* “HDMPFOSXXO” partially deciphers to “\_LAINTEXT”, the most likely English word is “PLAINTEXT”, then for “H”-> ”P”
* “BFGW” partially deciphers to “\_NOW”, the most likely English word is “KNOW”, then for “B”-> “K”
* “CKSJQSFODY” partially deciphers to “FRE\_UENTLY”, the most likely English word is “FREQUENTLY”, then for “J”-> ”Q”
* “LYEAGDL” partially deciphers to “SYM\_OLS”, the most likely English word is “SYMBOLS”, then for “A”->” B”

At this point, we have deciphered our entire ciphered text, and are left with the following key

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CIPHER | A | B | C | D | E | F | G | H | I | J | K | L | M |
| PLAIN | B | K | F | L | M | N | O | P | H | Q | R | S | A |
| CIPHER | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
| PLAIN | C | T | I | U | G | E |  | D | V | W | X | Y |  |

“T” and “Z” doesn’t appear in our cipher, and the remaining letter in the English alphabet are “Z” and “J”. In the English language, both “J” and “Z” are two of the least appearing letters. “Z” appears 0.27% of the time and “J” appears 0.19% of the time. As in the cipher both “T” and “Z” don’t make appearances, we could map either “Z” or “J” to “T” or “Z”. Until we receive more cipher text where either of the two appear.

The reference list consists of the online tools or resources we used to help us decipher the text.

# References

(2023). Retrieved from Calculate letter Frequencies: https://onlinetoolz.net/letter-frequency#text=GFS%20WMY%20OG%20LGDVS%20MF%20SFNKYHOSU%20ESLLMRS,%20PC%20WS%0ABFGW%20POL%20DMFRQMRS,%20PL%20OG%20CPFU%20M%20UPCCSKSFO%0AHDMPFOSXO%20GC%20OIS%20LMES%20DMFRQMRS%20DGFR%20SFGQRI%20OG%0ACPDD%20GFS%20LISSO%20GK%20

*Cryptanalysis Hints*. (2023). Retrieved from https://www3.nd.edu/~busiforc/handouts/cryptography/cryptography%20hints.html#:~:text=The%20most%20common%20two%2Dletter,words%20are%20the%20and%20and.

*The frequency of the letters of the alphabet in English*. (2023). Retrieved from https://www3.nd.edu/~busiforc/handouts/cryptography/letterfrequencies.html